Accelerating antibiotic discovery by leveraging machine learning models: application to identify novel inorganic complexes

Miroslava Nedyalkova[1], Gozde Demirci[1], Youri Cortat[1], Kevin Schindler[1], Fatlinda Rhamani[1], Aurelien Crochet[1], Aleksandar Pavic[2], Fabio Zobi[1*], and Marco Lattuada[1*]

[1] Department of Chemistry, Fribourg University, Chemin Du Musée 9, 1700 Fribourg, Switzerland
[2] Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Vojvode Stepe 444a, 11042 Belgrade, Serbia
[*] marco.lattuada@unifr.ch; fabio.zobi@unifr.ch

Abstract

The expanded prevalence of resistant bacteria and the inherent challenges of complicated infections highlight the urgent need to develop credible antibiotic options. Through conventional screening approaches, the discovery of new antibiotics has proven to be challenging. Anti-infective drugs, including antibacterials, antivirals, antifungals, and antiparasitics, have become less effective due to the spread of drug resistance. In this work we help define the design of next-generation antibiotic analogs based on metal complexes. The primary direction is based on the application of artificial intelligence (AI) methods, which demonstrated superior ability in tackling resistance in Gram-positive and Gram-negative bacteria, including multidrug-resistant strains. The bottleneck of the existing AI approaches relies on the structure similarities of the current antibiotics. The question of discovering and developing new unconventional antibiotic classes has challenged preconceptions about the scope and applicability of the existing methods. Herein, we developed a machine learning approach that predicts the minimum inhibitory concentration (MIC) of Re-complexes towards two *S. aureus* strains (ATCC 43300 - MRSA and ATCC 25923 - MSSA). Multi-layer Perceptron (MLP) was tailored with the structure features of the Re-complexes to develop the prediction model. Although our approach is demonstrated with a specific example, based on the rhenium carbonyl complexes, the predictive model can be readily adjusted to other candidate metal complexes. The model emphasizes applying a developed approach in the de novo design of a metal-based antibiotic with targeted activity against a challenging pathogen.

Introduction

It has been recognized that the emergence of antibiotic-resistant microbes represents a "clear and present danger" with a global impact, and, therefore, an effective response should be facilitated by adopting a novel approach for the de-novo design of novel classes of new antibiotics candidates. It has become increasingly demanding to treat infections that threaten global health due to the emergence and spread of pathogens that are resistant to conventional antibiotics. Some of the problematic pathogens are carbapenem-resistant *Enterobacteriaceae* (CRE), methicillin-resistant *Staphylococcus aureus* (MRSA), multidrug-resistant tuberculosis (MDR-TB), vancomycin-resistant *Enterococcus* (VRE), extended-spectrum beta-lactamase (ESBL)–producing bacteria, and drug-resistant *Candida auris*, *Neisseria gonorrhoeae*, *P. falciparum*, and *Toxoplasma gondii*. Challenges include antimicrobial stewardship (the appropriate and responsible use of

anti-infective drugs), developing new anti-infective drugs, increasing existing drugs against resistant infections, and understanding the drug mechanism. The UN 2023 report 'Bracing for Superbugs: Strengthening Environmental Action in the One Health Response to Antimicrobial Resistance,' which predicts that deaths resulting from drug-resistant infections will increase dramatically by 2050. Moreover, the problem could potentially lead to a $3.4 trillion shortfall in GDP over the next decade and push 24 million more people into extreme poverty unless drastic action is taken to address it. Approximately 1.27 million deaths worldwide were directly related to drug-resistant infections in 2019, while 4.95 million deaths were directly related to bacterial antimicrobial resistance (AMR). For comparison, in the same year, 860,000 deaths were attributed to HIV/AIDS and 640,000 to malaria, respectively, according to the World Health Organization (WHO) statistics.

How does resistance develop? Resistance develops due to genetic modification, mutations, or transmission of antibiotic-resistant genes from other microbes, unnecessary prescriptions, or the wrong combination of antibiotics[1,2]. Drug resistance has led to a decrease in the effectiveness of anti-infective drugs, such as antibacterials, antivirals, antifungals, and antiparasitics. An additional issue that must be addressed is the development of superbugs which are strains of bacteria that are resistant to several types of antibiotics. There is a lack of fast and accurate detection of infections, as well as an increase in antimicrobial resistance, which exacerbates these problems. To combat infectious diseases, new anti-infective therapies are urgently required, especially those that offer novel chemical spaces or therapeutic modalities. A major application of AI, and in particular machine learning (ML), a subfield of AI that uses data to train machines to make predictions, has been the facilitation of searches of small molecule databases, such as ZINC15[3]. The application of machine learning to the discovery of anti-infective drugs has centered on training models to identify potential new drugs or new uses for existing drugs. Since the number of drug-like small molecules is virtually infinite, 10^{60} and possibly greater[4], most antibiotics are not considered drug-like[5]. ML approaches provide a major advantage in that they enable the screening of compound libraries, and chemical spaces of drug-like compounds. The development of the first antibiotic designed by a computer with proven efficacy in preclinical animal models demonstrates that machines and AI could be used to develop therapeutic molecules[6]. The new avenues for ML-facilitated antibiotic discovery are envisioned to trickle in from algorithmic theory, AI, and molecular design. Creative integration of diverse studies from DL, computer-aided drug design, and ML research will play essential roles in accelerating the urgent task of novel antibiotic discovery and implementing new concepts towards the desired pharmacological activity discovery[7,8].

Conventional organic medicinal chemistry has faced difficulties and obstacles to replace the depleted antimicrobial pipeline. The data based on new antibiotics discoveries from 2022 revealed that, as of June 2021, there were only 45 ‘traditional’ antibiotics in clinical development. Hence, new approaches for designing the next generation of antibiotics are urgently ought. Inorganic or organometallic metal complexes have a pivotal role as an option for new antimicrobial alternatives. However, only recently have metals and metalloantibiotics gained considerable attention as potential antimicrobials, in response to the rapid rise of AMR in the past decade. Organometallic compounds hold promise thanks to the flexibility of their chemistry that allows one to change their structure and the nature of their ligands[9,10]. Amongst such species, rhenium-based complexes[11-17] hold great potential. Their mechanism of action is not fully understood, but current evidence points to the bacterial membranes as the target of compounds[18-20]. One of the first detailed mechanism-of-action investigations for metalloantibiotics was undertaken by the groups of Bandow and Metzler-Nolte[21]. They revealed that the trimetallic complex containing rhenium, iron and manganese with a peptide nucleic acid backbone showed good activity against a range of Gram-positive bacteria including MRSA, vancomycin-intermediate *S. aureus* and *Bacillus subtilis*. Unfortunately, the reported results about the activity against the tested Gram-negative pathogens were not with the same

measured response. In addition to in-depth mechanistic studies, a structure–activity relationship was carried out, which demonstrated that the Re-containing [(dpa)Re(CO)$_3$] moiety was crucial for the activity responses, and ferrocenyl and CpMn(CO)$_3$ units could be replaced by non-metal moieties, such as a phenyl ring[11].

The discovery of new antimicrobial agents is a pivotal necessity, and the abilities of metal-based complexes as drugs, more specifically, as antibiotics should be explored, and we should address the challenges outlined earlier, about how to boost the development in the field. It is critically important to prescribe appropriate antimicrobial therapy as quickly as possible. Whole genome sequencing approaches for quickly identifying pathogens and predicting antibiotic resistance phenotypes are becoming increasingly feasible. They may effectively reduce turnaround times for clinical tests relative to traditional culture-based methods, thereby improving patient outcomes. Using whole genome sequencing data from 1668 clinical isolates of *Klebsiella pneumoniae*, Nguyen *et al.* developed a machine learning model based on XGBoost that accurately predicted antibiotic MICs[22]. The obtained MICs predicted by the model correlate with recognized antimicrobial resistance genes. Pataki *et al.* used 704 *Escherichia coli* genomes combined with measured MIC for ciprofloxacin collected from different countries for MIC prediction model based on Random Forest[23]. The model was developed to identify the genomic features that determine susceptibility. The recent progress in whole genome sequencing technology in combination with machine learning analysis approaches indicates that soon such an approach might become cheaper and faster than a MIC measurement[24-26].

As we have seen, in the last years ML could deliver an alternative approach to streamline the development process of *de novo* antibiotics by identifying the key motif in the molecular structure associated with antibiotic activity. The application of ML to drug discovery, specifically antibiotic discovery, has been greatly facilitated by the public availability of empirical datasets[27-30], such as the Open Antimicrobial Drug Discovery (CO-ADD) for metal-containing compounds with antimicrobial activity[9,31]. Antibacterial screening approaches still do not have efficient tools and strategies for rapidly increasing the number of new chemotypes. Implementing the ML methods for novel compounds acting against Gram-negative bacteria is scarcely used[32,33]. An ML-guided approach based on descriptor space search and selection has already been used to predict antimicrobial activity[34-37]. The way of representing molecules is a crucial step. Numeric vectors consisting of molecular descriptor values (features) have already been utilized before the widespread applicability of ML in QSAR (quantitative structure−activity relationship) modeling. Molecular graph networks are another way of describing molecules where a node represents each atom, and an edge represents each atom's bond. There is a close relationship between ML models and molecular representations. Depending on how the chemical structures are presented, the chemical space can be defined as the union of all possible chemical compounds that can be used to find new antibiotic candidates that resemble existing ones.

Herein, we present an application of the ML approach for the prediction of the antibacterial activity of new antibiotic candidates based on Re-metal complexes towards two *S. aureus* strains (namely ATCC 43300, MRSA, and ATCC 25923, MSSA), using molecular descriptors and neural network architectures. Our framework provides a rapid method for developing a model to predict the metal complexes' minimum inhibitory concentrations (MIC). It consists of the following elements: (1) molecular representation based on the structure of the complexes, (2) feature reduction space, (3) ML algorithm, and (4) molecular descriptor specificity analysis (featuring importance scores). By leveraging the physicochemical properties captured by the molecular structures of 119 Re-complexes, measured data points (minimum inhibitory concentration) towards the activity of *S. aureus* ATCC 43300 and *S. aureus* ATCC 25923 strains were used for the antimicrobial activity prediction of previously untested complexes.

Experimental section

The Multi-layer Perceptron (MLP) was used for the neural network model. MLP is a supervised learning algorithm, the most straightforward feed-forward network, as shown in Figure 5. In the architecture of the feed-forward neural networks, the units (or nodes) are arranged into a graph without any sequential loops. This contrasts with recurrent neural networks[38], where the graph can have loops, so the network feeds into itself from the loops. The MLP learns a function by training on a dataset and, consequently, the number of dimensions of input and the number of dimensions of output. It is different from logistic regression in that, between the input and the output layer, there can be one or more non-linear layers, called hidden layers. For the training of the model, MLP regressor trains using backpropagation with no activation function in the output layer, which can also be seen as using the identity function as an activation function. The model was trained on the provided data set having 119 data rows (input objects), 92 feature columns, and 2 output classes. This set was subdivided into 75% for the training set (89 data rows) and 25% for the validation set (30 data rows). The code used for the model, along with the training and test set, is provided in this repository: https://github.com/mici345/MIC-prediction-model

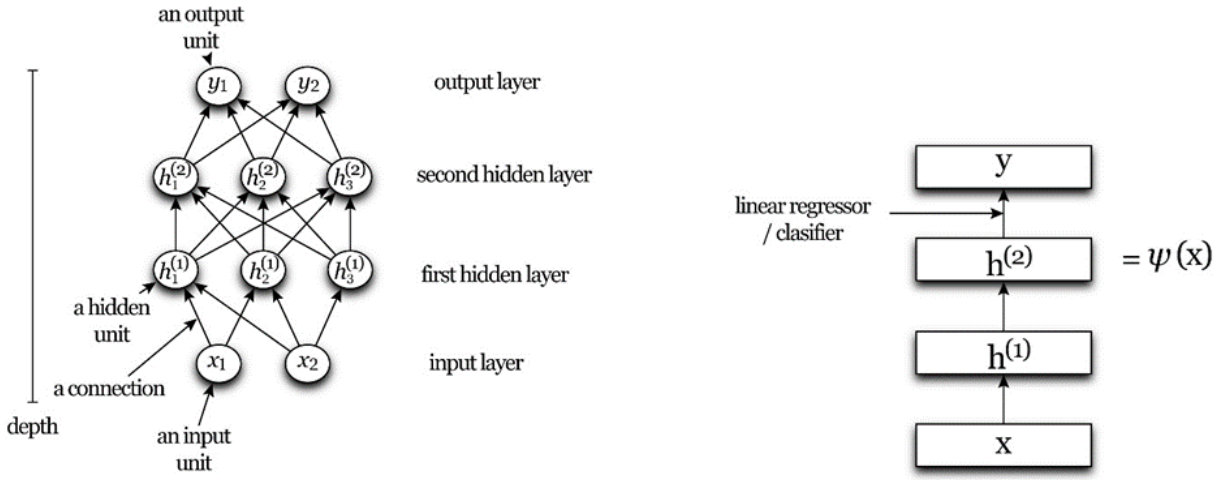

Figure 5. Multilayer perceptron with two hidden layers. Left picture: input layer, input feature values are used for the input units. The output layer has one unit per each value of the network outputs. Hidden layers: the layers between in and out units. Right picture: layers presented as boxes.

Structure of the dataset, descriptors generation space, and descriptors importance. We initially used 5666 descriptors for building the model, which are representations of Re-compounds. The AlvaDesc software was used to generate descriptors space from the 3D structures of each Re-complex. The set of used descriptors includes 0D (with no relation to shape, e.g., molecular weight), 1D (e.g., presence of certain active substructures within the molecule), 2D (e.g., molecular graph representations involving bonds between atoms but not bond lengths), and 3D (e.g., distances between specific atomic pairs in the molecule) ones (details in the Supporting Information). The descriptors contain information that could correlate to a given Re-complex's antimicrobial action. The structure of the input matrix for the ML models often leads to decreasing predictive accuracy. The reduction techniques are usually performed to decrease the noise in

the data structure, but at the same time, the loss of information should be presumed. The data sets were reduced with Principal Component Analysis (PCA) for reducing the descriptor space. PCA is an orthogonal linear transformation that transforms the data into a new coordinate system where the first direction of the most significant variance becomes the new coordinate axis[39]. The optimal parameter selection within the descriptors space resulted in highly converged accuracies for the trained model. Construction of the initial matrix from the explored chemical database is crucial for developing and validating the model. The final reduced set was based on 119 data points (Re-complex), 91 features, and 2 output classes for the bacterial strains.

Reagents and chemicals. All reagents were obtained from standard sources and utilized without any further purifications. The compounds $Re(CO)_5Br$ and $Re(CO)_5Cl$ was purchased from Sigma-Aldrich. For the validation set, complexes **1a**[40], **1b**[41], **1c**[42], **2a**[40], **2b**[43], **2c**[44], **3b**[45], **4b**[46], **7a**[40] and **7b**[47] were synthesized according to published procedures. Complexes **5d**, **6d**, **8d**, and **9d** were prepared according to Cortat *et al*.[17] Complexes **3a**, **3c**, **4a**, **4c**, **8a**, and **9a** were prepared with similar procedures.[17] All complexes were synthesized under an inert (Ar) atmosphere.

Instruments and Analysis. IR spectra were recorded on a Bruker TENSOR II with the following parameters: 16 scans for the background and 32 scans for the sample with a resolution of 4 cm^{-1} in the 4000 to 600 cm^{-1} regions. UV-Vis spectra of the complexes were measured on a Jasco V730 spectrophotometer. A Bruker Advance III 400 MHz was used to measure the complexes' NMR spectra. The corresponding 1H chemical shifts were reported relative to the residual solvent protons. A Bruker FTMS 4.7-T Apex II in positive mode was used to perform the mass analyses.

Synthetic procedures. Ligands for complexes **2a-c** and **7a-b** were synthesized using published procedures[48-51]. For the preparation of the complexes in the validation set, $Re(CO)_5Br$ and $Re(CO)_5Cl$ were used. Rhenium precursors and ligands were generally reacted in equimolar ratios and refluxed overnight. After the reactions, products were filtered and washed with the reaction solvent and diethyl ether. The purity of the complexes (Br or Cl species **3a**, **4a**, **8a**, and **9a**) was confirmed as >95%. Compounds **3c** and **4c** were prepared by suspending **3a** or **4a** in MeOH (HPLC grade) with 1-mole equiv. of pyridine and AgOTf (1.2 mole equiv.) and refluxing in the dark overnight. After the mixture had cooled to room temperature, it was filtered to discard AgBr and dried in a vacuum oven. The compounds were purified by washing with water, followed by centrifugation. Crystals of **4c** suitable for X-ray diffraction analysis were grown by slow evaporation of a dichloromethane solution at room temperature. Spectrochemical characterization of the complexes is in the Supporting Information. Crystallographic data, CCDC number 2296107, have been deposited at the Cambridge Crystallographic Data Centre.

Antimicrobial study. The antimicrobial activity of the complexes was assessed against *S. aureus* ATCC 25923 (wild type, MSSA) and *S. aureus* ATCC 43300 (methicillin-resistant, MRSA) strains following published protocols[52]. Briefly, each complex was prepared as a 6.4 mM stock solution in DMSO and diluted to 256 μM with PBS. They were sterilized for 20 min under UV light before use. Then, stock samples were diluted with PBS to 128 μM, and 50 μL of each dilution was transferred to 96-well plates. In parallel, *S. aureus* in Mueller–Hinton Broth (non-cation-adjusted, MHB), cultured one day before injection, was used to prepare bacterial suspensions at 1 x 106 CFU/mL in MHB 2X. Then, 50 μL of *S. aureus* suspensions were mixed in the 50 μL of serially diluted sample wells, leading to a final bacterial concentration of 5 x 105 and the complexes concentrations ranging from 64 μM to 0.5 μM. The plates were incubated at 37 °C for 24 h. The minimum inhibitory concentration (MIC) values were determined by measuring the optical density at 600 nm (OD600). The assay was conducted in triplicate. Tecan-Spark 10M with SparkControl program was used to determine the antimicrobial activities.

Results and Discussion

Synthesis of metal complexes. A dataset of 119 rhenium tricarbonyl complexes was compiled for this work. Of these, 20 complexes, not previously evaluated for their antibiotic activity, were synthesized and used as the validation set for the model (Figure 1). Complexes shown in Figure 1 were prepared according to well-established procedures employed in the chemistry of this metal core. $[Re(CO)_3(NN)X]$ species (where NN = diimine ligand and X = halide; Br or Cl) were obtained by reacting commercially available $Re(CO)_5X$ with one equivalent of NN in refluxing toluene. The desired compounds precipitate upon cooling and are filtered and washed with a cold solvent to yield the molecules with a purity $\geqslant$ 95%. Pyridine (py) and clotrimazole (ctz) derivatives of the compounds were prepared by reaction of $[Re(CO)_3(NN)X]$ with AgOTf in the presence of py or ctz, followed by precipitation and HPLC purification if required. Characterization of the new compounds is given in SI. In addition to standard characterization techniques, the x-ray structure of **4c** was determined (SI).

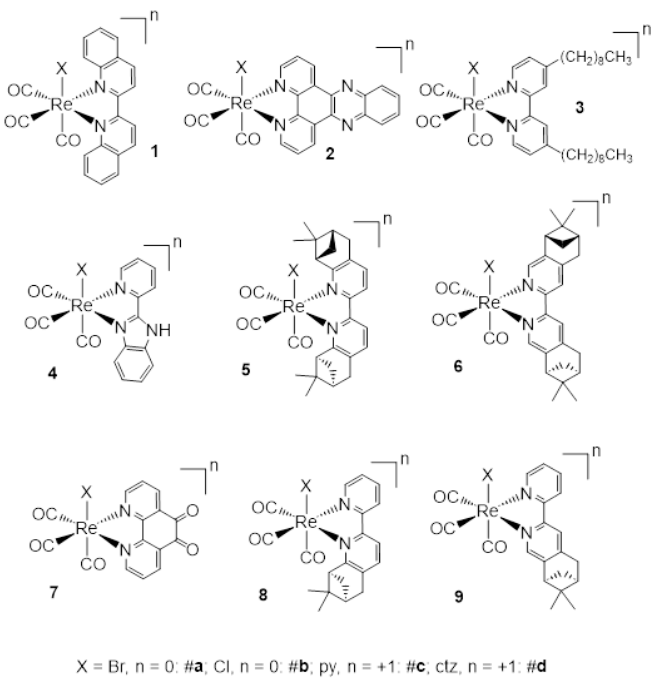

Figure 1. Structures of validation complexes tested for antimicrobial activity against *S. aureus* ATCC 25923 (MSSA) and *S. aureus* ATCC 43300 (MRSA) strains.

Multi-layer Perceptron (MLP) model. The pipeline for leveraging the prediction model based on antimicrobial data is presented in Figure 2. Using the MLP, we predicted the antimicrobial activity, quantified in the MIC values Re-complexes. We trained the MLP architecture on the whole set. The obtained scores were computed based on the validation set and showed how the model performs by comparing the model's predictions to the experimental validation values. The prediction performance was evaluated using the following parameters: accuracy, precision, and recall. Accuracy estimates how often the model predictions were correct. As such, it is the ratio of the true cases to all the cases, defined as (TP

+ TN) / (TP + FP + TN + FN) where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, FN is the number of false negatives. The set of labels predicted for a sample must match the corresponding labels in the validation set. Precision indicates how often the model predicted the sample to be positive when true. It is defined as the ratio of the True Positive to the predicted positive cases. The precision is equal to TP / (TP + FP). It is intuitively the ability of the classifier not to label as positive a sample that is negative. Recall quantifies the number of positive predictions made from all positive cases in the dataset, equal to TP / (TP + FN). The classifier intuitively can find all the positive samples. The obtained scores are shown in Table 1 and Figure 3.

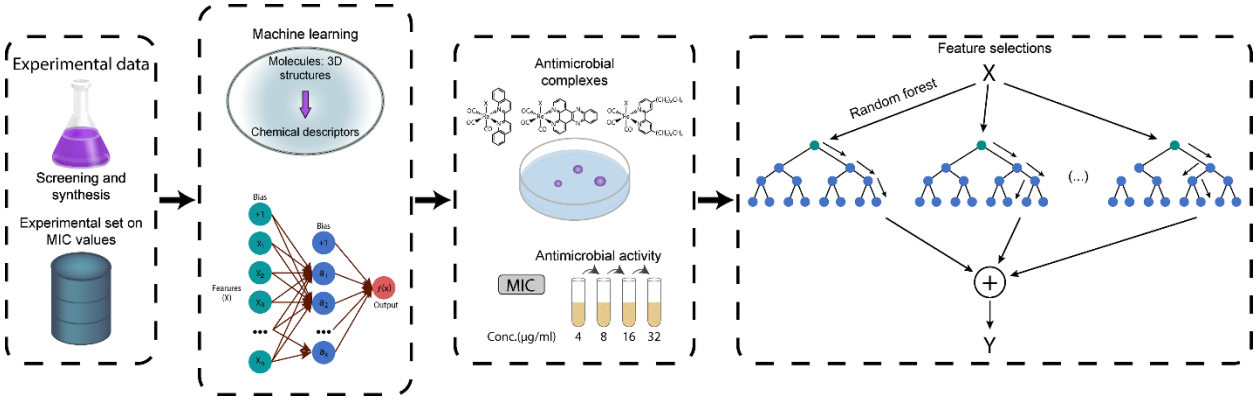

Figure 2. Schematic representation of the workflow. We started from experimentally obtained data to compile the input matrix for training and testing the model. For validating the model, the new synthesized Re-complexes were used. The descriptors were extracted from the AlvaDesc software, and the MLP was used to predict the antimicrobial activity. The last box shows the scoring of the molecular features by recursive feature elimination based on bootstrap-aggregated decision trees.

Table 1. Model prediction performance evaluated by Accuracy, Precision, and Recall criteria.

Target	Accuracy	Precision	Recall
MRSA[a] – MIC [b] 4	0.9	0.85	0.92
MRSA - MIC 8	0.83	0.78	0.93
MRSA - MIC 16	0.87	0.81	1
MRSA - MIC 32	0.9	0.94	0.89
MSSA[c] - MIC 4	0.8	0.64	0.9
MSSA - MIC 8	0.83	0.75	0.92
MSSA - MIC 16	0.87	0.84	0.94
MSSA - MIC 32	0.73	0.74	0.82

[a] MRSA refers to the *S. aureus* ATCC 43300 strain. [b] MIC # indicates the target minimum inhibitory concentration in mM. [c] MSSA refers to the *S. aureus* ATCC 25923 strain.

Figure 3. Model prediction performance evaluated by Accuracy, Precision, and Recall criteria. For all these scores, the best value is 1, and the worst is 0.

In practical terms, with overall good accuracy, the model can predict whether a metal complex may be active against the MRSA *S. aureus* strain. Indeed, a good agreement between expected and experimental MIC values is found when considering the methicillin-resistant *S. aureus* ATCC 43300. The complete statistical metrics are presented in Table S4 in the SI section for the validation set. In this case, in 14 out of 20 instances (Table 2), the model correctly predicted whether a metal complex showed no activity (MIC > 32 μg/mL, 7 out of 10 experimentally inactive compounds) or if a complex had potential antibacterial activity (MIC < 32 μg/mL, 7 out of 10 experimentally active compounds). In this latter case, in 6 out of 7 instances, the MIC value was correctly predicted within 2x of the experimentally determined MIC value. The model revealed more limitations when considering the wild-type *S. aureus* ATCC 25923 (MSSA) strain. However, in the case of experimentally active molecules, not only did the model always correctly identify these complexes (8 out of 8 instances within 2x the experimentally determined MIC value), but in 3 cases, it was able to indicate the correct MIC value of the compounds (**1c**, **4a** and **4c**, Figure 4).

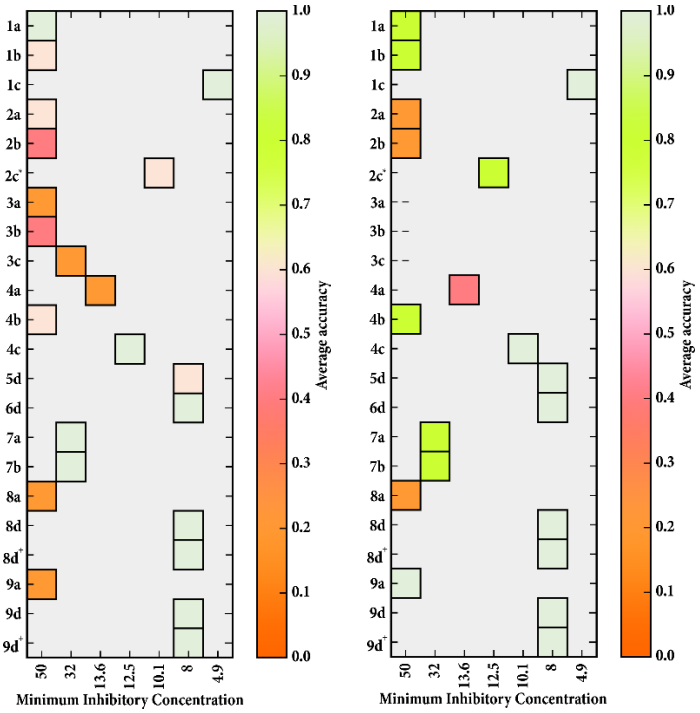

Figure 4. The accuracy of the model for the MIC values (μg/mL) of the tested complexes against MRSA and MSSA *S. aureus* strains.

Feature ranking analysis. The effect of features and network architectures on the quality of the predictions was tested, and the distribution of the final scores for the features is presented in Figure 5. After evaluating the model performance, we examined the feature importance in the prediction model. The Neural Network is not a straightforward method to assess the intrinsic feature importance. For example, it is hard to interpret how these weights contribute to the resulting decisions just by analyzing the weights between the neurons in the model. The feature ranking of the set of used descriptors (presented as an additional file in the SI) was calculated based on their relative contributions to predictions made by the model. Below, we try to assess the significance of feature j as follows:

$$f_j = s - \frac{1}{n}\sum_{i=1}^{n} s_{ij}$$

where s – is the baseline score computed for the non-permuted input data set; basically, it is equal to 1; s_{ij} – is the score obtained by permuting the corresponding feature column of the data set. By permuting, we mean that the values of the feature are randomly permuted between various data rows (molecules). In this way, the importance of a feature is the difference between the baseline score *s* and the average score obtained by permuting the corresponding column of the test set. If the difference is small, then the model is insensitive to permutations of the feature, so its importance is low. On the contrary, the features' importance is high if the difference is significant. The parameter n controls the number of permutations per feature — more permutations yield better estimates (we used n=100).

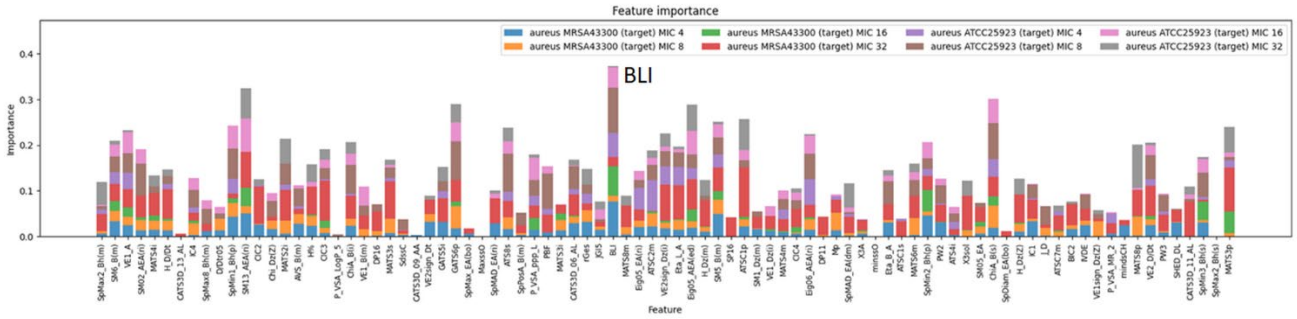

Figure 5. Molecular descriptor scoring.

This assessment with the low scores does not necessarily mean that the feature is not important at all but rather that most of its values are close to each other and possibly that when training. The top descriptor with the highest average effect in the prediction model is listed in Table 3. The Kier benzene-likeliness index (BLI) descriptors, which are calculated by dividing the first-order valence connectivity index by the number of non-H bonds (nBO) of the molecule and then normalizing on the benzene molecule proposed to measure the molecule aromaticity, were defined as a top-ranked feature of importance. This is true for all cases except the last case (*S. aureus* ATCC25923 (target) MIC 32).

Table 3. Top-scored descriptors for each of the target cases.

Target	Categories of descriptors	Description
S. aureus MRSA MIC 4	BLI - Topological indices[53]	Kier benzene-likeliness index is an aromaticity index calculated from molecular topology.
S. aureus MRSA MIC 8	BLI - Topological indices[53]	
S. aureus MRSA MIC 16	BLI - Topological indices[53]	
S. aureus MRSA MIC 32	BLI - Topological indices[53]	
S. aureus MSSA MIC 4	BLI - Topological indices[53]	
S. aureus MSSA MIC 8	BLI - Topological indices[53]	
S. aureus MSSA MIC 16	BLI - Topological indices[53]	
S. aureus MSSA MIC 32	Edge adjacency indices[54]	Spectral moment of order 13 from augmented edge adjacency mat. weighted by resonance integral (structural properties of the graph)

A second method for the feature assessment was applied to investigate the effect and robustness of the obtained BLI descriptor as the top ranked. The ranking was performed using Random Forest (RF, Bootstrap-aggregated (bagged) decision trees)[55], trained on a random forest of 200 classification trees, and stored the out-of-bag information for predictor importance estimation. The critical values are sorted and reported in the Excel file (S2 in the SI) in the SI. BLI aromaticity index was the top-ranked. The relationship between the antimicrobial activity and the impact of the BLI, indicates that the BLI is a crucial descriptor in the

prediction models. The effect of the BLI descriptor could be connected to the lipophilicity of the molecules. Other features of high relevance are edge adjacency indices (e.g., eigenvalue or spectral mean absolute deviation indices such as Eig05_EA (ri), Eig05_AEA (ed) or SpMAD_EA (dm) and SM13_AEA (ri)).

The correlation between aromaticity, hydrophobicity and antimicrobial activity is well known for the antimicrobial peptides[56-58] and was also shown for antimicrobial polymers[59]. In this last study, the logarithm of the partition coefficient of compounds between n-octanol (C logP) and water was used to represent hydrophobicity. The authors showed that the elevated responses from the antimicrobial activity required hydrophobicity that was neither too high nor too low. The obtained values suggest that C logP values between 0 and 2 have the best balance of high antimicrobial activity. Then, the obtained relations between the identified descriptors in the MLP model were explainable. The nature of the target should be tested as a limiting factor for the sparser application of the obtained results about the activity towards different bacterial cells. Still, it might display highly synergistic effects based on the selectivity of the descriptor.

Conclusion

This study presents a model based on supervised learning methods to predict the antibacterial properties of Re-metals complexes. This machine-learning-guided descriptor model was developed on Re-metals complexes and was proven able to predict the antimicrobial activity of the metal complexes against the methicillin-resistant *S. aureus* ATCC 43300 strain with good accuracy and precision. It may thus serve as an advisory tool to guide the synthesis of new complexes. The MLP model makes use of (1) molecular representation based on the structure of the complexes, (2) feature reduction space, (3) ML algorithm, and (4) molecular descriptor specificity analysis (features *importance* scores). When applied to the prediction of 20 previously untested molecules, in 70% of cases, it could predict whether (a) the metal complex may be active or (b) inactive. Moreover, in >80% of correctly predicted active molecules, their minimum inhibitory concentration was predicted within 2x the experimentally determined values. The proposed ML-based antibiotic development approach revealed the main descriptors responsible for the antimicrobial activity. Therefore, the model may predict the activity of new unconventional antibacterial candidates based on Re-complexes with the selected molecular descriptors.

Data availability

The code used for the model, along with the training and test set, is provided in this repository: https://github.com/mici345/MIC-prediction-model

Competing interests

There are no competing financial and non-financial interests to declare.

Author contributions

M.N., F.Z., and M.L. designed the project. M.N. performed the computational study. G.D. performed screening antimicrobial tests and synthesized validation complexes. Y.C .and, K.S. and F.R. performed the synthesis of training set complexes. A.C. collected crystallographic data and solved the structure of **4c**. A.P. performed antimicrobial testing. M.N., F.Z., and M.L. original draft, writing-review& editing.

References

1 Maillard, J. Y. Resistance of Bacteria to Biocides. *Microbiol. Spectr.* **6**, doi:10.1128/microbiolspec.ARBA-0006-2017 (2018).

2 Andersson, D. I. & Hughes, D. Selection and Transmission of Antibiotic-Resistant Bacteria. *Microbiol. Spectr.* **5**, doi:10.1128/microbiolspec.MTBP-0013-2016 (2017).

3 Sterling, T. & Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **55**, 2324-2337, doi:10.1021/acs.jcim.5b00559 (2015).

4 Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97-113, doi:10.1038/nrd.2017.232 (2018).

5 O'Shea, R. & Moser, H. E. Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery. *J. Med. Chem.* **51**, 2871-2878, doi:10.1021/jm700967e (2008).

6 Torres, M. D. T. & de la Fuente-Nunez, C. Toward computer-made artificial antibiotics. *Curr. Opin. Microbiol.* **51**, 30-38, doi:https://doi.org/10.1016/j.mib.2019.03.004 (2019).

7 Torres, M. D. T., Sothiselvam, S., Lu, T. K. & de la Fuente-Nunez, C. Peptide Design Principles for Antimicrobial Applications. *J. Mol. Biol.* **431**, 3547-3567, doi:https://doi.org/10.1016/j.jmb.2018.12.015 (2019).

8 de la Fuente-Nunez, C. Toward Autonomous Antibiotic Discovery. *mSystems* **4**, doi:10.1128/mSystems.00151-19 (2019).

9 Frei, A. *et al.* Metal complexes as a promising source for new antibiotics. *Chem. Sci.* **11**, 2627-2639, doi:10.1039/C9SC06460E (2020).

10 Frei, A., Verderosa, A. D., Elliott, A. G., Zuegg, J. & Blaskovich, M. A. T. Metals to combat antimicrobial resistance. *Nat. Rev. Chem.* **7**, 202-224, doi:10.1038/s41570-023-00463-4 (2023).

11 Patra, M. *et al.* An organometallic structure-activity relationship study reveals the essential role of a Re(CO)3 moiety in the activity against gram-positive pathogens including MRSA. *Chem. Sci.* **6**, 214-224, doi:10.1039/C4SC02709D (2015).

12 Siegmund, D. *et al.* Benzannulated Re(I)–NHC complexes: synthesis, photophysical properties and antimicrobial activity. *Dalton Trans.* **46**, 15269-15279, doi:10.1039/C7DT02874A (2017).

13 Slate, A. J., Shalamanova, L., Akhidime, I. D. & Whitehead, K. A. Rhenium and yttrium ions as antimicrobial agents against multidrug resistant Klebsiella pneumoniae and Acinetobacter baumannii biofilms. *Lett. Appl. Microbiol.* **69**, 168-174, doi:https://doi.org/10.1111/lam.13154 (2019).

14 Sovari, S. N. *et al.* Design, synthesis and in vivo evaluation of 3-arylcoumarin derivatives of rhenium(I) tricarbonyl complexes as potent antibacterial agents against methicillin-resistant Staphylococcus aureus (MRSA). *Eur. J. Med. Chem.* **205**, 112533, doi:https://doi.org/10.1016/j.ejmech.2020.112533 (2020).

15 Sovari, S. N. *et al.* Combatting AMR: A molecular approach to the discovery of potent and non-toxic rhenium complexes active against C. albicans-MRSA co-infection. *Eur. J. Med. Chem.* **226**, 113858, doi:https://doi.org/10.1016/j.ejmech.2021.113858 (2021).

16 Cooper, S. M. *et al.* Synthesis and anti-microbial activity of a new series of bis(diphosphine) rhenium(v) dioxo complexes. *Dalton Trans.* **51**, 12791-12795, doi:10.1039/D2DT02157A (2022).

17 Cortat, Y. *et al.* Computer-Aided Drug Design and Synthesis of Rhenium Clotrimazole Antimicrobial Agents. *Antibiotics* **12**, 619 (2023).

18 Wenzel, M. *et al.* Analysis of the Mechanism of Action of Potent Antibacterial Hetero-tri-organometallic Compounds: A Structurally New Class of Antibiotics. *ACS Chem. Biol.* **8**, 1442-1450, doi:10.1021/cb4000844 (2013).

19 Mendes, S. S. *et al.* Synergetic Antimicrobial Activity and Mechanism of Clotrimazole-Linked CO-Releasing Molecules. *ACS Bio & Med Chem Au* **2**, 419-436, doi:10.1021/acsbiomedchemau.2c00007 (2022).
20 Schindler, K. *et al.* Antimicrobial Activity of Rhenium Di- and Tricarbonyl Diimine Complexes: Insights on Membrane-Bound S. aureus Protein Binding. *Pharmaceuticals* **15**, 1107 (2022).
21 Wenzel, M. *et al.* Analysis of the mechanism of action of potent antibacterial hetero-tri-organometallic compounds: a structurally new class of antibiotics. *ACS Chem. Biol.* **8**, 1442-1450, doi:10.1021/cb4000844 (2013).
22 Nguyen, M. *et al.* Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. *Sci. Rep.* **8**, 421, doi:10.1038/s41598-017-18972-w (2018).
23 Pataki, B. Á. *et al.* Understanding and predicting ciprofloxacin minimum inhibitory concentration in Escherichia coli with machine learning. *Sci. Rep.* **10**, 15026, doi:10.1038/s41598-020-71693-5 (2020).
24 Jeukens, J. *et al.* Genomics of antibiotic-resistance prediction in Pseudomonas aeruginosa. *Ann. N.Y. Acad. Sci.* **1435**, 5-17, doi:https://doi.org/10.1111/nyas.13358 (2019).
25 Eyre, D. W. *et al.* WGS to predict antibiotic MICs for Neisseria gonorrhoeae. *J. Antimicrob. Chemother.* **72**, 1937-1947, doi:10.1093/jac/dkx067 (2017).
26 Avershina, E. *et al.* AMR-Diag: Neural network based genotype-to-phenotype prediction of resistance towards β-lactams in Escherichia coli and Klebsiella pneumoniae. *CSBJ* **19**, 1896-1906, doi:https://doi.org/10.1016/j.csbj.2021.03.027 (2021).
27 Ruiz Puentes, P. *et al.* Rational Discovery of Antimicrobial Peptides by Means of Artificial Intelligence. *Membranes* **12**, 708 (2022).
28 Ren, Y. *et al.* Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics* **38**, 325-334, doi:10.1093/bioinformatics/btab681 (2021).
29 Huang, Y. *et al.* High-throughput microbial culturomics using automation and machine learning. *Nat. Biotechnol.*, doi:10.1038/s41587-023-01674-2 (2023).
30 Skinnider, M. A. *et al.* Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* **11**, 6058, doi:10.1038/s41467-020-19986-1 (2020).
31 Frei, A. *et al.* Metal Complexes as Antifungals? From a Crowd-Sourced Compound Library to the First In Vivo Experiments. *JACS Au* **2**, 2277-2294, doi:10.1021/jacsau.2c00308 (2022).
32 Durrant, J. D. & Amaro, R. E. Machine-Learning Techniques Applied to Antibacterial Drug Discovery. *Chem. Biol. Drug Des.* **85**, 14-21, doi:https://doi.org/10.1111/cbdd.12423 (2015).
33 Martin, E. J. *et al.* All-Assay-Max2 pQSAR: Activity Predictions as Accurate as Four-Concentration IC50s for 8558 Novartis Assays. *J. Chem. Inf. Model.* **59**, 4450-4459, doi:10.1021/acs.jcim.9b00375 (2019).
34 Tiihonen, A. *et al.* Predicting Antimicrobial Activity of Conjugated Oligoelectrolyte Molecules via Machine Learning. *J. Am. Chem. Soc.* **143**, 18917-18931, doi:10.1021/jacs.1c05055 (2021).
35 Medvedeva, A., Teimouri, H. & Kolomeisky, A. B. Predicting Antimicrobial Activity for Untested Peptide-Based Drugs Using Collaborative Filtering and Link Prediction. *J. Chem. Inf. Model.* **63**, 3697-3704, doi:10.1021/acs.jcim.3c00137 (2023).
36 Ishfaq, M., Aamir, M., Ahmad, F., M Mebed, A. & Elshahat, S. Machine Learning-Assisted Prediction of the Biological Activity of Aromatase Inhibitors and Data Mining to Explore Similar Compounds. *ACS Omega* **7**, 48139-48149, doi:10.1021/acsomega.2c06174 (2022).
37 Diéguez-Santana, K. & González-Díaz, H. Machine learning in antibacterial discovery and development: A bibliometric and network analysis of research hotspots and trends. *Comput. Biol. Med.* **155**, 106638, doi:https://doi.org/10.1016/j.compbiomed.2023.106638 (2023).

38 Schmidt, R. M. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. *ArXiv* **abs/1912.05911**, doi:https://doi.org/10.48550/arXiv.1912.05911 (2019).
39 Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**, 115-129, doi:10.1007/BF02289694 (1964).
40 Kurz, P., Probst, B., Spingler, B. & Alberto, R. Ligand Variations in [ReX(diimine)(CO)3] Complexes: Effects on Photocatalytic CO2 Reduction. *Eur. J. Inorg. Chem.* **2006**, 2966-2974, doi:10.1002/ejic.200600166 (2006).
41 Machura, B., Kruszynski, R. & Kusz, J. X-ray structure, spectroscopic characterisation and DFT calculations of the [Re(CO)3(dppt)Cl] complex. *Polyhedron* **26**, 1590-1596, doi:https://doi.org/10.1016/j.poly.2006.11.034 (2007).
42 Moya, S. A. *et al.* Influence of the 4-Substituted Pyridine Ligand L′ on both the Conformation and Spectroscopic Properties of the (2,2′-Biquinoline-κN1,κN1′)tricarbonyl(pyridine-κN1)rhenium(1+) Complex ([Re(CO)3-(bqui)(py)]+) and Its Derivatives [Re(CO)3(L)(L′)]+ (L = 2,2′-Biquinoline and 3,3′-(Ethane-1,2-diyl)-2,2′-biquinoline). *Helv. Chim. Acta* **88**, 2842-2860, doi:https://doi.org/10.1002/hlca.200590227 (2005).
43 Ruiz, G. T. *et al.* Intercalation of fac-[(4,4′-bpy)ReI(CO)3(dppz)]+, dppz = dipyridyl[3,2-a:2′3′-c]phenazine, in polynucleotides. On the UV-vis photophysics of the Re(i) intercalator and the redox reactions with pulse radiolysis-generated radicals. *Dalton Trans.*, 2020-2029, doi:10.1039/B614970G (2007).
44 Wing-Wah Yam, V., Kam-Wing Lo, K., Cheung, K.-K. & Yuen-Chong Kong, R. Deoxyribonucleic acid binding and photocleavage studies of rhenium(I) dipyridophenazine complexes. *J. Chem. Soc., Dalton Trans.*, 2067-2072, doi:10.1039/A700828G (1997).
45 Klein, D. M. *et al.* Shorter Alkyl Chains Enhance Molecular Diffusion and Electron Transfer Kinetics between Photosensitisers and Catalysts in CO2-Reducing Photocatalytic Liposomes. *Chem. Eur. J.* **27**, 17203-17212, doi:https://doi.org/10.1002/chem.202102989 (2021).
46 Tzeng, B.-C. *et al.* pH-Dependent Spectroscopic and Luminescent Properties, and Metal-Ion Recognition Studies of Re(I) Complexes Containing 2-(2′-Pyridyl)benzimidazole and 2-(2′-Pyridyl)benzimidazolate. *Inorg. Chem.* **50**, 5379-5388, doi:10.1021/ic1019058 (2011).
47 Kaplanis, M. *et al.* Re(I) tricarbonyl complex of 1,10-phenanthroline-5,6-dione: DNA binding, cytotoxicity, anti-inflammatory and anti-coagulant effects towards platelet activating factor. *J. Inorg. Biochem.* **135**, 1-9, doi:https://doi.org/10.1016/j.jinorgbio.2014.02.003 (2014).
48 Molphy, Z. *et al.* Copper Phenanthrene Oxidative Chemical Nucleases. *Inorg. Chem.* **53**, 5392-5404, doi:10.1021/ic500914j (2014).
49 Greguric, A., Greguric, I. D., Hambley, T. W., Aldrich-Wright, J. R. & Collins, J. G. Minor groove intercalation of Δ-[Ru(Me2phen)2dppz]2+ to the hexanucleotide d(GTCGAC)2. *J. Chem. Soc., Dalton Trans.*, 849-855, doi:10.1039/B105689C (2002).
50 Wang, C. *et al.* Increasing the triplet lifetime and extending the ground-state absorption of biscyclometalated Ir(iii) complexes for reverse saturable absorption and photodynamic therapy applications. *Dalton Trans.* **45**, 16366-16378, doi:10.1039/C6DT02416E (2016).
51 Nagaraj, K., Senthil Murugan, K., Thangamuniyandi, P. & Sakthinathan, S. Synthesis, Micellization Behaviour, DNA/RNA Binding and Biological Studies of a Surfactant Cobalt(III) Complex With Dipyrido[3,2-a:2′,4′-c](6,7,8,9-tetrahydro)phenazine. *J. Fluoresc.* **24**, 1701-1714, doi:10.1007/s10895-014-1457-1 (2014).
52 Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163-175, doi:10.1038/nprot.2007.521 (2008).

53 Basak, S. C., Balaban, A. T., Grunwald, G. D. & Gute, B. D. Topological Indices: Their Nature and Mutual Relatedness. *J. Chem. Inf. Comput. Sci.* **40**, 891-898, doi:10.1021/ci990114y (2000).
54 Estrada, E., Guevara, N. & Gutman, I. Extension of Edge Connectivity Index. Relationships to Line Graph Indices and QSPR Applications. *J. Chem. Inf. Comput. Sci.* **38**, 428-431, doi:10.1021/ci970091s (1998).
55 Tin Kam, H. in *Proceedings of 3rd International Conference on Document Analysis and Recognition.* 278-282 vol.271.
56 Clark, S., Jowitt, T. A., Harris, L. K., Knight, C. G. & Dobson, C. B. The lexicon of antimicrobial peptides: a complete set of arginine and tryptophan sequences. *Commun. Biol.* **4**, 605, doi:10.1038/s42003-021-02137-7 (2021).
57 Jin, L. *et al.* A Designed Tryptophan- and Lysine/Arginine-Rich Antimicrobial Peptide with Therapeutic Potential for Clinical Antibiotic-Resistant Candida albicans Vaginitis. *J. Med. Chem.* **59**, 1791-1799, doi:10.1021/acs.jmedchem.5b01264 (2016).
58 Chan, D. I., Prenner, E. J. & Vogel, H. J. Tryptophan- and arginine-rich antimicrobial peptides: structures and mechanisms of action. *Biochim. Biophys. Acta* **1758**, 1184-1202, doi:10.1016/j.bbamem.2006.04.006 (2006).
59 Phuong, P. T. *et al.* Effect of Hydrophobic Groups on Antimicrobial and Hemolytic Activity: Developing a Predictive Tool for Ternary Antimicrobial Polymers. *Biomacromolecules* **21**, 5241-5255, doi:10.1021/acs.biomac.0c01320 (2020).